HIERARCHICAL GAUSSIAN FILTER

A Bayesian foundation for individual learning under uncertainty

Lucile Alys Favero

1$^{\text{er}}$ novembre 2022

PLAN

AIM. Explain the HIERARCHICAL GAUSSIAN FILTER model

1. Definition of Hierarchical Gaussian Filter model
2. Definition of the invert problem
3. Variational inversion
    1. Assumptions
    2. Reformulation of the problem
    3. Maximization of negative free energy with Lagrange multipliers
4. Quadratic approximation to the variational energy
5. Update equations
6. Simulations

1. Definition of Hierarchical Gaussian Filter model

1. Definition of Hierarchical Gaussian Filter model

Describe how an agent learns about a continuous uncertain quantity (ie random variable) x that moves.

- We can describe this motion with a Gaussian random walk :

$$x^{(k)} \sim \mathcal{N}\left(x^{(k-1)}, \theta\right), k = 1, 2, ... \tag{1}$$

  where k is the time index and θ is a constant positive.
  But there is no reason to assume that the volatility in x is constant.

- Replace θ by a positive function f of a second random variable x_2 , while x become x_1 :

$$x_1^{(k)} \sim \mathcal{N}\left(x_1^{(k-1)}, f(x_2)\right), k = 1, 2, ...$$

- May assume x_2 performs a Gaussian random walk of its own with a constant variance θ. Same as (1).

1. Definition of Hierarchical Gaussian Filter model

- Can continue adding levels of Gaussian random walks coupled by their variances up to any number n.

$$x_i^{(k)} \sim \mathcal{N}\left(x_i^{(k-1)}, f_i(x_i)\right), i = 1, 2, ...n-1$$

  At the top level, instead of $f_n(x_{n+1}$, we have θ a constant

- At each level i, the coupling to the next hightest level $i+1$, is given by :

$$f_i(x_{i+1}) := t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}$$

  where :
  - $\{\kappa_1, \omega_1, ..., \kappa_{n-1}, \omega_{n-1}, \vartheta\} =: \chi$ model parameters
  - $t^{(k)}$ allows input coming at irregular intervals

- Observations are denoted as $u^{(k)}$

1. Definition of Hierarchical Gaussian Filter model

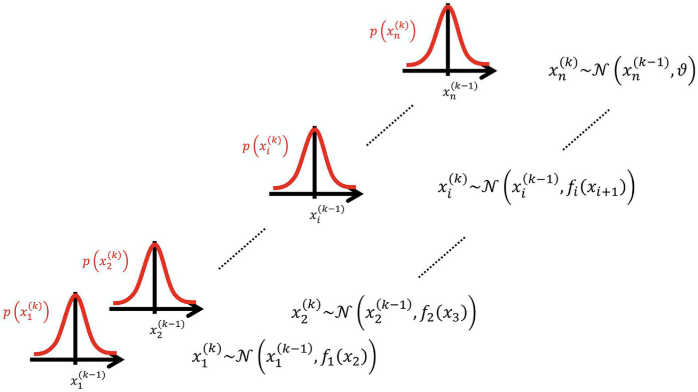

FIGURE 1 | Overview of the Hierarchical Gaussian Filter (HGF). The model represents a hierarchy of coupled Gaussian random walks. The levels of the hierarchy relate to each other by determining the step size (volatility or variance) of a random walk. The topmost step size is a constant parameter ϑ.

1. Definition of Hierarchical Gaussian Filter model

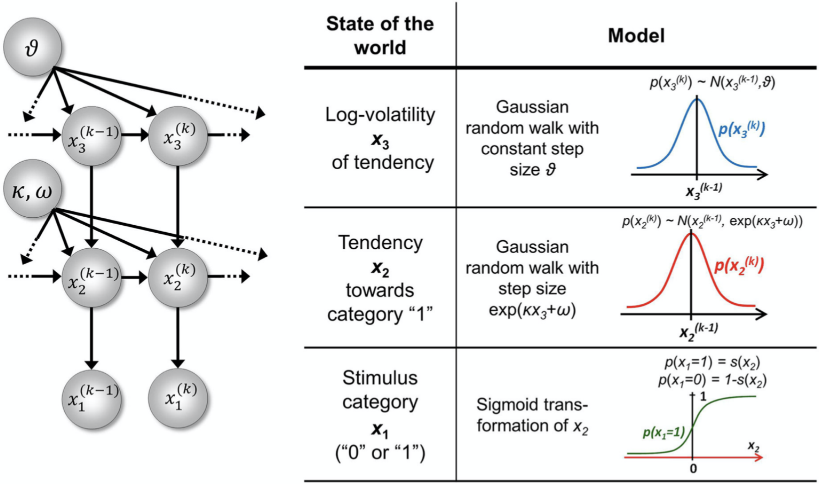

FIGURE 2 | The 3-level HGF for binary outcomes. The lowest level, x_1, is binary and corresponds, in the absence of sensory noise, to sensory input u. Left: schematic representation of the generative model as a Bayesian network. $x_1^{(k)}$, $x_2^{(k)}$, $x_3^{(k)}$ are hidden states of the environment at time point k. They generate $u^{(k)}$, the input at time point k, and depend on their immediately preceding values $x_2^{(k-1)}$, $x_3^{(k-1)}$ and on the on parameters κ, ω, ϑ. Right: model definition. This figure has been adapted from Figures 1, 2 in Mathys et al. (2011).

2. Definition of the invert problem

2. Definition of the invert problem : Simplification of the problem

The inversion problem is to find the posterior $p\big(x^{(k)}, \chi | u^{(1...k)}\big)$. According to the formula for conditional probability, we have :

$$p\big(x^{(k)}, \chi | u^{(1...k)}\big) = p\big(x^{(k)} | \chi, u^{(1...k)}\big) p\big(\chi | u^{(1...k)}\big) \tag{2}$$

We considere that $p\big(\chi | u^{(1...k)}\big) = \delta(\chi - \chi_a)$, where χ_a are the fixed parameter values that characterize a particular agent at a particular time. So it remains to compute $p\big(x^{(k)} | \chi, u^{(1...k)}\big)$.
The problem can be formulated as follow :
Knowing the structure of the environment - ie the law followed by each x_i -, the realisation up to time k and the parameters χ, we want to know $x^{(k)}$

2. Definition of the invert problem : problematic of the exact inversion

According to Bayes' theorem :

$$p\big(x^{(k)}|\chi, u^{(1...k)}\big) = \frac{p\big(\chi, u^{(1...k)}|x^{(k)}\big)p\big(x^{(k)}\big)}{p\big(\chi, u^{(1...k)}\big)} = \frac{p\big(\chi, u^{(1...k)}|x^{(k)}\big)p\big(x^{(k)}\big)}{\color{red}{\int p\big(\chi, u^{(1...k)}, x^{(k)}\big)dx^{(k)}}} \tag{3}$$

To compute the red expression we need to perform a marginalization over $x^{(k)}$. But this is intractable.

3. Variational inversion

 First assumption. Variational assumption

The posterior distribution is approximated by a **variational distribution** :

$$p\big(x^{(k)}|\chi, u^{(1...k)}\big) \approx q\big(x^{(k)}|\chi, u^{(1...k)}\big)$$

The distribution q is restricted to belong to a family of distributions of simpler form than p and selected with the intention of making q similar to p. Here we will choose :

$$q\big(x^{(k)}|\chi, u^{(1...k)}\big) = \mathcal{N}\Big(x^{(k)}, \mu^{(k)}, \sigma^{(k)}\Big)$$

Second assumption. Mean field approximation

$q\big(x^{(k)}|\chi, u^{(1...k)}\big)$ is fully factorized over the hidden variables :

$$q\big(x^{(k)}|\chi, u^{(1...k)}\big) = \prod_{i=1}^{n} q_i\big(x_i^{(k)}|\chi, u^{(1...k)}\big) \tag{4}$$

where $x = \Big(x^{(1)}, \ldots, x^{(k)}\Big)$ and $x^{(j)} = \Big(x_1^{(j)}, \ldots, x_n^{(j)}\Big)$ (There are k steps of times and n levels of hidden variables)

2. Reformulation of the problem

- Thanks to theses assumptions, the problem of finding the posterior p, became **finding** $q_i, i \in \{1, ..., n\}$ **the more similar to** p**.**
- **Kullback–Leibler divergence** $D_{KL}\big(Q||P\big) := \sum_Z Q\big(Z\big) \ln \dfrac{Q\big(Z\big)}{P\big(Z|X\big)}$

  The Kullback–Leibler divergence (KL-divergence) of P from Q can be see as the choice of dissimilarity function. So we can reformulate our problem as **the minimization of** $D_{KL}\big(q||p\big)$
- **The negative free energy** $\mathcal{F}(Q) := -\sum_Z Q\big(Z\big)\big(\ln Q\big(Z\big) - \ln P(Z, X)\big)$

  We can show that min $D_{KL}\big(Q||P\big)$ is equivalent to max $\mathcal{F}(Q)$. So the problem become **max** $\mathcal{F}(q)$

Show that minimization of $D_{KL}\big(q||p\big)$ is equivalent to max $\mathcal{F}(Q)$.

$$
\begin{aligned}
D_{KL}\big(Q||P\big) &:= \sum_Z Q\big(Z\big) \ln \frac{Q\big(Z\big)}{P\big(Z|X\big)} \\
&= \sum_Z Q\big(Z\big) \ln \Big(\frac{Q\big(Z\big)P\big(X\big)}{P\big(Z,X\big)}\Big) \\
&= \underbrace{\sum_Z Q\big(Z\big)\Big(\ln Q\big(Z\big) - \ln P\big(Z,X\big)\Big)}_{\mathcal{F}(Q)} + \underbrace{\sum_Z Q\big(Z\big) \ln P\big(X\big)}_{\ln P\big(X\big)}
\end{aligned}
$$

So,

$$
\ln P\big(X\big) = D_{KL}\big(Q||P\big) + \mathcal{F}(Q)
$$

As $\ln P(X)$ is fixed with respect to Q maximizing $\mathcal{F}(Q)$ minimizes the KL divergence of Q from P

Notation : we definie $q_{x^{(k)}}\left(x^{(k)}\right) := q\left(x^{(k)} | \chi, u^{(1...k)}\right)$:
Compute the negative free energy $\mathcal{F}$

$$\begin{aligned}
\mathcal{F}\left(q_{x^{(k)}}\left(x^{(k)}\right)\right) &= \int q_{x^{(k)}}\left(x^{(k)}\right) \ln\left(\frac{p(x^{(k)}, \chi | u^{(1...k)})}{q_{x^{(k)}}\left(x^{(k)}\right)} dx^{(k)}\right) && \text{by definition} \\
&= \int q_{x^{(k)}}\left(x^{(k)}\right) \ln\left(p(x^{(k)}, \chi | u^{(1...k)})\right) dx^{(k)} \\
&\quad - \int q_{x^{(k)}}\left(x^{(k)}\right) \ln\left(q_{x^{(k)}}\left(x^{(k)}\right)\right) dx^{(k)} && \text{by property of the logarithm} \\
&= \int \prod_{i=1}^{n} q_{x_i^{(k)}}\left(x_i^{(k)}\right) \ln\left(p(x^{(k)}, \chi | u^{(1...k)})\right) dx^{(k)} \\
&\quad \underbrace{- \int \prod_{i=1}^{n} q_{x_i^{(k)}}\left(x_i^{(k)}\right) \ln\left(\prod_{i=1}^{n} q_{x_i^{(k)}}\left(x_i^{(k)}\right)\right) dx^{(k)}}_{=: - \sum_{j=1}^{n} q_{x_j^{(i)}}\left(x_j^{(i)}\right) \ln\left(q_{x_j^{(i)}}\left(x_j^{(i)}\right)\right)} && \text{by (4), mean field assumption}
\end{aligned}$$

 3. Maximization of $\mathcal{F}$

We use Lagrange multipliers : $\{\lambda_i\}$ with $i \in \{1,..,n\}$ on $\mathcal{F}$, with the constraints of normalisation :
$\int q_{x_i^{(k)}}(x_i^{(k)})dx_i^{(k)} = 1$.
We have the new functional :

$$\tilde{\mathcal{F}}\left(q_{x^{(k)}}(x^{(k)})\right) = \mathcal{F}\left(q_{x^{(k)}}(x^{(k)})\right) + \sum_{i=1}^{n} \lambda_i \Big(\int q_{x_i^{(k)}}(x_i^{(k)})dx_i^{(k)} - 1\Big)$$

We then take the functional derivative of this expression with respect to each $q_{x_i^{(k)}}(x_i^{(k)})$, and equalize to zero.
Let's fix $i \in \{1,..,k\}$ we have :

$$\begin{aligned}\frac{\partial \tilde{\mathcal{F}} q_{x_j^{(i)}}(x_j^{(i)})}{\partial q_{x_j^{(i)}}(x_j^{(i)})} &= \frac{\partial \mathcal{F} q_{x_j^{(i)}}(x_j^{(i)})}{\partial q_{x_j^{(i)}}(x_j^{(i)})} + \lambda_{ij} \\ &= \int \prod_{\substack{h=1 \\ h\neq j}}^{n} q_{x_h^{(i)}}(x_h^{(i)}) \ln\left(p(x^{(i)}, \chi | u^{(1...k)})\right) dx_{\setminus j}^{(i)} - \ln\left(q_{x_j^{(i)}}(x_j^{(i)})\right) + \lambda_{ij} = 0\end{aligned}$$

Where $dx_{\setminus j}^{(i)} := \{dx_h^{(i)}\}_{h \neq j}$.

Isolating $q_{x_j^{(i)}}\left(x_j^{(i)}\right)$:

$$\begin{aligned} \ln\left(q_{x_j^{(i)}}\left(x_j^{(i)}\right)\right) &= \int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\setminus j}^{(i)} + \lambda_{ij} \\ \Longrightarrow q_{x_j^{(i)}}\left(x_j^{(i)}\right) &= e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\setminus j}^{(i)}} e^{\lambda_{ij}} \end{aligned} \tag{5}$$

We inject this expression of $q_{x_j^{(i)}}\left(x_j^{(i)}\right)$ in the constraints of normalisation equation $\int q_{x_j^{(i)}}\left(x_j^{(i)}\right)dx_j^{(i)} = 1$ and isolate λ_{ij} :

$$\begin{aligned}
&\int q_{x_j^{(i)}}\left(x_j^{(i)}\right)dx_j^{(i)} = 1 \\
\Longrightarrow &\int e^{\lambda_{ij}} e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\backslash j}^{(i)}} dx_j^{(i)} = 1 \\
\Longrightarrow & e^{\lambda_{ij}} \int e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\backslash j}^{(i)}} dx_j^{(i)} = 1 \\
\Longrightarrow & e^{\lambda_{ij}} = \left(\int e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\backslash j}^{(i)}} dx_j^{(i)}\right)^{-1}
\end{aligned} \tag{6}$$

Now we substitute this expression into the one of $q_{x_j^{(i)}}\left(x_j^{(i)}\right)$ (5)

$$q_{x_j^{(i)}}\left(x_j^{(i)}\right) = e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\setminus j}^{(i)}} e^{\lambda_{ij}} \qquad \text{by } (5)$$

$$= e^{\underbrace{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\setminus j}^{(i)}}_{=:I\left(x_j^{(i)}\right)}} \Bigg(\underbrace{\int e^{\int \prod_{\substack{h=1\\h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)},\chi|u^{(1...k)})\right) dx_{\setminus j}^{(i)}} dx_j^{(i)}}_{=:Z_j^{(i)}}\Bigg)^{-1} \qquad \text{by } (6)$$

So we have :

$$q_{x_j^{(i)}}\left(x_j^{(i)}\right) = \frac{1}{Z_j^{(i)}} e^{I\left(x_j^{(i)}\right)} \tag{7}$$

With :

$$Z_j^{(i)} = \int e^{\int \prod_{\substack{h=1 \\ h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)}, \chi | u^{(1...k)})\right) dx_{\setminus j}^{(i)}} dx_j^{(i)}$$

and

$$I\left(x_j^{(i)}\right) = \int \prod_{\substack{h=1 \\ h\neq j}}^{n} q_{x_h^{(i)}}\left(x_h^{(i)}\right) \ln\left(p(x^{(i)}, \chi | u^{(1...k)})\right) dx_{\setminus j}^{(i)} \tag{8}$$

4.Quadratic approximation to the variational energy

4. Quadratic approximation to the variational energy

We will want to find a quadratic approximation of I. Indeed we will use this quadratic property to compute the update equation of μ and σ (cf next section)
To do so we will use a power series approximation up to second order

1. Compute $p(x^{(k)}, \chi | u^{(1...k)})$
2. Compute I
3. Power series

4.Quadratic approximation

For this computation we will use :

1. Definition of $q\left(x_{\setminus i}^{(k)}\right)$

$$q\left(x_{\setminus i}^{(k)}\right) = \prod_{\substack{j=1 \\ j \neq i}}^{n} q\left(x_j^{(k)}\right) \tag{9}$$

2. Property of distribution of probability

$$\int q\left(x_i^{(k)}\right) dx_i^{(k)} = 1 \tag{10}$$

3. Definition of the normal distribution, (where here σ is the variance)

$$\mathcal{N}(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}} \tag{11}$$

4. Definition of the mean

$$\int q\left(x_i^{(k)}\right) x_i^{(k)} dx_i^{(k)} = \mu_i^{(k)} \tag{12}$$

5. Definition of the variance

$$\int q\left(x_i^{(k)}\right) \left(x_i^{(k)} - \mu_i^{(k)}\right)^2 dx_i^{(k)} = \sigma_i^{(k)} \tag{13}$$

4.Quadratic approximation. **Compute** $p(x^{(k)}, \chi | u^{(1...k)})$

Factorization
Let fix i, we definie : $x_{\setminus i} = \{x_j\}_{j \neq i}$.
Then, we have :

$$\begin{aligned}
p\left(x^{(k)}, \chi | u^{(1...k)}\right) &= p\left(x_i^{(k)}, x_{\setminus i}^{(k)}, \chi | u^{(1...k)}\right) \\
&= \int_{x_1^{(k-1)}} \cdots \int_{x_n^{(k-1)}} \prod_{j=1}^{n} p\left(x_j^{(k)} | x_j^{(k-1)}, x_{\setminus j}^{(k)}, \chi, u^{(1...k)}\right) \\
&\quad p\left(x_j^{(k-1)} | u^{(1...k-1)}\right) dx_1^{(k-1)} \ldots dx_n^{(k-1)} \\
&= \prod_{j=1}^{n} \int_{x_j^{(k-1)}} p\left(x_j^{(k)} | x_j^{(k-1)}, x_{\setminus j}^{(k)}, \chi, u^{(1...k)}\right) p\left(x_j^{(k-1)} | u^{(1...k-1)}\right) dx_j^{(k-1)}
\end{aligned}$$

where :

- the first equality by the notation definied
- the second equality by the formula of the total probability
- in the third equality we factorize the integral since each term i depends only on $x_i^{(k-1)}$.

4.Quadratic approximation. **Compute** $p(x^{(k)}, \chi | u^{(1...k)})$

Compute $\int p\left(x_j^{(k)} | x_j^{(k-1)}, x_{\backslash j}^{(k)}, \chi\right) p\left(x_j^{(k-1)} | u^{(1...k-1)}\right) dx_j^{(k-1)}$ for $j \in \{1...n\}$ fixed

- **By definition of the model HFG**, the next inference at level j, time k, knowing the previous inferences and all the other level at time k folllow a normal walk centred in the inference at the previous time, same level, and with a variance of $t^{(k)} e^{\kappa_j x_{j+1}^{(k)} + \omega_j}$, (see appendice A Mathys 2014) for the justification)

$$p\left(x_j^{(k)} | x_j^{(k-1)}, x_{\backslash j}^{(k)}, \chi\right) = \mathcal{N}\left(x_j^{(k)}, x_j^{(k-1)}, t^{(k)} e^{\kappa_j x_{j+1}^{(k)} + \omega_j}\right)$$

  Where for $j = n$, the variance is considered as constante : $t^{(k)} e^{\kappa_n x_{n+1}^{(k)} + \omega_n} = \theta$

- **By assumption of the form of the posterior.** We did the assumption that the posterior at each level j and time $k-1$ follows a normal distribution centred on $\mu_j^{(k-1)}$, and with a variance of $\sigma_j^{(k-1)}$.

$$p\left(x_j^{(k-1)} | u^{(1...k-1)}\right) = \mathcal{N}\left(x_j^{(k-1)}, \mu_j^{(k-1)}, \sigma_j^{(k-1)}\right)$$

So, we obtain :

$$
\begin{aligned}
&\int_{x_j^{(k-1)}} p\Big(x_j^{(k)}|x_j^{(k-1)},x_{\backslash j}^{(k)},\chi\Big)p\Big(x_j^{(k-1)}|u^{(1...k-1)}\Big)dx_j^{(k-1)}\\
&=\int_{x_j^{(k-1)}} \mathcal{N}\Big(x_j^{(k)},x_j^{(k-1)},t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\Big)\mathcal{N}\Big(x_j^{(k-1)},\mu_j^{(k-1)},\sigma_j^{(k-1)}\Big)dx_j^{(k-1)}\\
&=\mathcal{N}\Big(x_j^{(k)},\mu_j^{(k-1)},\sigma_j^{(k-1)}+t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\Big)
\end{aligned}
$$

Where the ultimate equality come from the following property, proved in appendix :

$$
\int_{\mu}\mathcal{N}\Big(z|\mu,\sigma_0^2\Big)\mathcal{N}\Big(\mu|\mu_0,\sigma_1^2\Big)d\mu=\mathcal{N}\Big(z|\mu_0,\sigma_0^2+\sigma_1^2\Big) \tag{14}
$$

Compute $p\big(x^{(k)},\chi|u^{(1...k)}\big)$
So we can deduce :

$$
p\big(x^{(k)},\chi|u^{(1...k)}\big)=\prod_{j=1}^{n}\mathcal{N}\Big(x_j^{(k)},\mu_j^{(k-1)},\sigma_j^{(k-1)}+t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\Big) \tag{15}
$$

4.Quadratic approximation **Compute** I

We will now use the expression of $p\left(x^{(k)},\chi|u^{(1...k)}\right)$ to compute I.

$$I(x_i^{(k)}) = \int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(p\left(x^{(k)},\chi|u^{(1...k)}\right)\right)dx_{\backslash i}^{(k)} \qquad \text{by } (8) : \text{definition of } I$$

$$= \int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(\prod_{j=1}^{n}\mathcal{N}\left(x_j^{(k)},\mu_j^{(k-1)},\sigma_j^{(k-1)}+t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\right)\right)dx_{\backslash i}^{(k)} \qquad \text{by } (15)$$

$$= \sum_{j=1}^{n}\int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(\mathcal{N}\left(x_j^{(k)},\mu_j^{(k-1)},\sigma_j^{(k-1)}+t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\right)\right)dx_{\backslash i}^{(k)} \qquad \text{by property of the logarith}$$

$$= \sum_{\substack{j=1\\ j\neq i-1\\ j\neq i}}^{n}\underbrace{\int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(\mathcal{N}\left(x_j^{(k)},\mu_j^{(k-1)},\sigma_j^{(k-1)}+t^{(k)}e^{\kappa_j x_{j+1}^{(k)}+\omega_j}\right)\right)dx_{\backslash i}^{(k)}}_{\text{don't depend on } x_i^{(k)} \text{ so it is a constant}} \qquad \longleftarrow \text{ others terms}$$

$$+\underbrace{\int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(\mathcal{N}\left(x_{i-1}^{(k)},\mu_{i-1}^{(k-1)},\sigma_{i-1}^{(k-1)}+t^{(k)}e^{\kappa_{i-1} x_{i}^{(k)}+\omega_{i-1}}\right)\right)dx_{\backslash i}^{(k)}}_{=-\ln\pi-\frac{1}{2}\ln\left(\sigma_{i-1}^{(k-1)}+t^{(k)}e^{\kappa_{i-1}x_i^{(k)}+\omega_{i-1}}\right)-\frac{1}{2}\frac{\sigma_{i-1}^{(k-1)}+\left(\mu_{i-1}^{(k)}-\mu_{i-1}^{(k-1)}\right)^2}{\sigma_{i-1}^{(k)}+t^{(k)}e^{\kappa_{i-1}x_i^{(k)}+\omega_{i-1}}}} \qquad \longleftarrow \text{ term } j=i-1$$

$$+\underbrace{\int\dots\int q\left(x_{\backslash i}^{(k)}\right)\ln\left(\mathcal{N}\left(x_{i}^{(k)},\mu_{i}^{(k-1)},\sigma_{i}^{(k-1)}+t^{(k)}e^{\kappa_{i} x_{i+1}^{(k)}+\omega_{i}}\right)\right)dx_{\backslash i}^{(k)}}_{=:\star\star} \qquad \longleftarrow \text{ term } j=i$$

We will now compute the integrales of the term $j = i$ in the expression of $I(x_i^{(k)})$. Using $x_{\setminus i} = \{x_j\}_{j \neq i}$ and (9) in a similar way as above, we have :

$$\star\star = \int_{x_n^{(k)}} \cdots \int_{x_1^{(k)}} q\left(x_{\setminus i}^{(k)}\right) \ln\left(\mathcal{N}\left(x_i^{(k)}, \mu_i^{(k-1)}, \sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right) dx_1^{(k)} \dots dx_{i-1}^{(k)} dx_{i+1}^{(k)} \dots dx_n^{(k)}$$

$$= \int_{x_n^{(k)}} \cdots \int_{x_1^{(k)}} \prod_{\substack{j=1 \\ j \neq i}}^{n} q\left(x_j^{(k)}\right) \ln\left(\mathcal{N}\left(x_i^{(k)}, \mu_i^{(k-1)}, \sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right) dx_1^{(k)} \dots dx_{i-1}^{(k)} dx_{i+1}^{(k)} \dots dx_n^{(k)}$$

The argument of the integral depend only on $x_i^{(k)}$ and $x_{i+1}^{(k)}$ we can put this term outside of the integrals on $dx_j^{(k)}$,with $j \neq i+1$, and obtain :

$$\star\star = \int_{x_n^{(k)}} \cdots \int_{x_2^{(k)}} \prod_{\substack{j=2 \\ j \neq i}}^{n} q\left(x_j^{(k)}\right) \ln\left(\mathcal{N}\left(x_i^{(k)}, \mu_i^{(k-1)}, \sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right) \underbrace{\int_{x_1^{(k)}} q\left(x_1^{(k)}\right) dx_1^{(k)}}_{=1 \text{ by } (10)} \dots dx_{i+1}^{(k)} dx_{i+1}^{(k)} \dots$$

$$= \dots$$

$$= \int q\left(x_{i+1}^{(k)}\right) \ln\left(\mathcal{N}\left(x_i^{(k)}, \mu_i^{(k-1)}, \sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right) dx_{i+1}^{(k)}$$

Applying the definition of the normal distribution, we separe the terms depending on $x_i^{(k)}$ and those not :

$$
\begin{aligned}
\star\star &= \int q\left(x_{i+1}^{(k)}\right) \ln\left(\frac{1}{\sqrt{2\pi\left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)}} e^{-\frac{1}{2}\frac{\left(x_{i+1}^{(k)} - \mu_i^{(k-1)}\right)^2}{\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_i^{(k)} + \omega_i}}} \right) dx_{i+1}^{(k)} \\
&= \int q\left(x_{i+1}^{(k)}\right) \left(-\frac{1}{2} \ln\left(2\pi\left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right) - \frac{1}{2} \frac{\left(x_{i+1}^{(k)} - \mu_i^{(k-1)}\right)^2}{\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_i^{(k)} + \omega_i}} \right) dx_{i+1}^{(k)} \\
&= \underbrace{\int q\left(x_{i+1}^{(k)}\right) \left(-\frac{1}{2} \ln\left(2\pi\left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_{i+1}^{(k)} + \omega_i}\right)\right)\right) dx_{i+1}^{(k)}}_{=\text{constante because do not depend on } x_i^{(k)}} \\
&\qquad - \frac{1}{2}\left(x_{i+1}^{(k)} - \mu_i^{(k-1)}\right)^2 \int q\left(x_{i+1}^{(k)}\right) \left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_i^{(k)} + \omega_i}\right)^{-1} dx_{i+1}^{(k)}
\end{aligned}
\tag{16}
$$

We develop $\left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i x_i^{(k)} + \omega_i}\right)^{-1}$ with a Taylor serie around $\mu_{i+1}^{(k-1)}$ up to the second order, and obtain :

$$\star\star = cst - \frac{1}{2}\big(x_{i+1}^{(k)} - \mu_i^{(k-1)}\big)^2 \left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i \mu_{i+1}^{(k-1)} + \omega_i}\right)^{-1}$$

So :

$$\begin{aligned} I(x_i^{(k)}) = cst - \frac{1}{2} \ln \left(\sigma_{i-1}^{(k-1)} + t^{(k)} e^{\kappa_{i-1} x_i^{(k)} + \omega_{i-1}}\right) - \frac{1}{2} \frac{\sigma_{i-1}^{(k-1)} + \big(\mu_{i-1}^{(k)} - \mu_{i-1}^{(k-1)}\big)^2}{\sigma_{i-1}^{(k)} + t^{(k)} e^{\kappa_{i-1} x_i^{(k)}}} \\ + \omega_{i-1} - \frac{1}{2}\big(x_i^{(k)} - \mu_i^{(k-1)}\big)^2 \left(\sigma_i^{(k-1)} + t^{(k)} e^{\kappa_i \mu_{i+1}^{(k-1)} + \omega_i}\right)^{-1} \end{aligned}$$

It is a quadratic function.

 Notations

$$
\begin{aligned}
A &:= \sigma_{i-1}^{(k)} + \left(\mu_{i-1}^{(k)} - \mu_{i-1}^{(k-1)}\right)^2 \\
v_i^{(k)} &:= \begin{cases} t^{(k)} e^{\kappa_i \mu_{i+1}^{(k-1)} + \omega_i}, & i = 1, ..., n-1 \\ t^{(k)} \vartheta, & i = n \end{cases} \\
\hat{\mu}_i^{(k)} &:= \mu_i^{(k-1)} \\
\hat{\pi}_i^{(k)} &:= \frac{1}{\sigma_i^{(k-1)} + v_i^{(k)}} \\
\delta_i^{(k)} &:= \left(\sigma_i^{(k)} + \left(\mu_i^{(k)} - \hat{\mu}_i^{(k)}\right)^2\right)\hat{\pi}_i^{(k)} - 1
\end{aligned}
$$

$$
I(x_i^{(k)}) = -\frac{1}{2}\left(\ln\left(\sigma_{i-1}^{(k-1)} + t^{(k)} e^{\kappa_{i-1} x_i^{(k)} + \omega_{i-1}}\right) + \frac{A}{\sigma_{i-1}^{(k-1)} + t^{(k)} e^{\kappa_{i-1} x_i^{(k)} + \omega_{i-1}}} + \hat{\pi}_i^{(k)} \left(x_i^{(k)} - \mu_i^{(k-1)}\right)^2 \right)
$$

5. Updates equations

5. Updates equations of $\sigma_i^{(k)}$

By previous computation we have :

$$\begin{cases} q_{x_i^{(k)}}\left(x_i^{(k)}\right) = \frac{1}{Z_i^{(k)}} e^{I\left(x_i^{(k)}\right)} & \text{by } (7)\text{, the mean field approximation} \\ q_{x_i^{(k)}}\left(x_i^{(k)}\right) = \frac{1}{\sqrt{2\pi\sigma_i^{(k)}}} e^{-\frac{(x_i^{(k)}-\mu_i^{(k)})^2}{2\sigma_i^{(k)}}} & \text{by the assumption on the gaussian form of } q \end{cases}$$

So :

$$q_{x_i^{(k)}}\left(x_i^{(k)}\right) = \frac{1}{\sqrt{2\pi\sigma_i^{(k)}}} e^{-\frac{(x_i^{(k)}-\mu_i^{(k)})^2}{2\sigma_i^{(k)}}} = \frac{1}{Z_i^{(k)}} e^{I\left(x_i^{(k)}\right)} \tag{17}$$

By taking the logarithm on both sides we have :

$$-2\ln\left(2\pi\sigma_i^{(k)}\right) - \frac{\left(x_i^{(k)}-\mu_i^{(k)}\right)^2}{2\sigma_i^{(k)}} = -\ln\left(Z^{(k)}\right) + I\left(x_i^{(k)}\right)$$

Differentiating twice with respect to $x_i^{(k)}$ gives :

$$-\frac{1}{\sigma_i^{(k)}} = I''\left(x_i^{(k)}\right) \tag{18}$$

5. Update equations of $\sigma_i^{(k)}$

Since I is a quadratic function, I'' is constant. In order to find it we may evaluate it in a known point as $\mu_i^{(k-1)}$.
The update equation of $\sigma_i^{(k)}$ is :

$$\pi_i^{(k)} = -I"(\mu_i^{(k-1)}) = \frac{1}{2}\left(\kappa_{i-1} v_{i-1}^{(k)} \hat{\pi}_{i-1}^{(k)}\right)^2 \left(1 + \delta_{i-1}^{(k)}\left(1 - \frac{1}{v_{i-1}^{(k)} \pi_{i-1}^{(k-1)}}\right)\right) + \hat{\pi}_i^{(k)} \tag{19}$$

Where $\pi_i^{(k)} = \frac{1}{\sigma_i^{(k)}}$

5. Update equations of $\mu_i^{(k)}$

- $\mu_i^{(k)}$ is the argument of the maximum of $I(x_i^{(k)})$
- Apply Newton method

  Starting at any point $x_i^{(k)}$ the exact argmax of a quadratic function can be found in one step by Newton's method :

$$\mu_i^{(k)} = \arg\max I\left(x_i^{(k)}\right) = x_i^{(k)} - \frac{I'\left(x_i^{(k)}\right)}{I''\left(x_i^{(k)}\right)}$$

We choose : $x_i^{(k)} = \mu_i^{(k)}$

$$\begin{aligned}
\mu_i^{(k)} &= \mu_i^{(k-1)} - \frac{I'\left(\mu_i^{(k-1)}\right)}{I''\left(\mu_i^{(k-1)}\right)} \\
&= \mu_i^{(k-1)} + \frac{1}{\pi_i^{(k)}} I'\left(\mu_i^{(k-1)}\right) \qquad (18)
\end{aligned}$$

And so the update equation of $\mu_i^{(k)}$ is :

$$\mu_i^{(k)} = \mu_i^{(k-1)} + \frac{1}{2}\kappa_{i-1} v_{i-1}^{(k)} \frac{\hat{\pi}_{i-1}^{(k)}}{\pi_{i-1}^{(k)}} \delta_{i-1}^{(k)}$$

6. Simulations

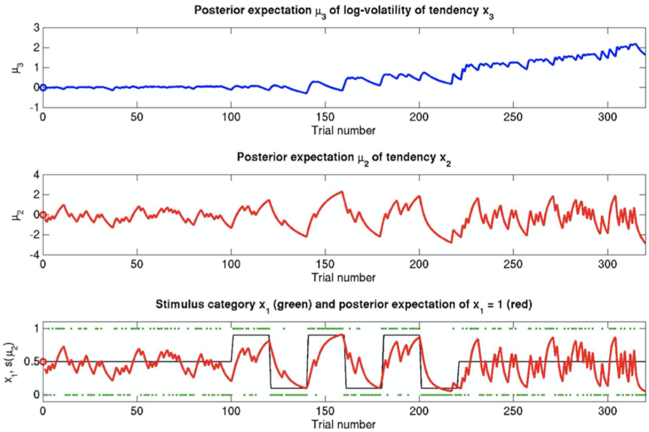

FIGURE 5 | Reference scenario: $\vartheta = 0.5$, $\omega = -2.2$, $\kappa = 1.4$. A simulation of 320 trials. Bottom: the first level. Input *u* is represented by green dots. In the absence of perceptual uncertainty, this corresponds to x_1. The fine black line is the true probability (unknown to the agent) that $x_1 = 1$. The red line shows $s(\mu_2)$; i.e., the agent's posterior expectation that $x_1 = 1$. Given the input and update rules, the simulation is uniquely determined by the value of the parameters ϑ, ω, and κ. Middle: the second level with the posterior expectation μ_2 of x_2. Top: the third level with the posterior expectation μ_3 of x_3. In all three panels, the initial values of the various μ are indicated by circles at trial $k = 0$.

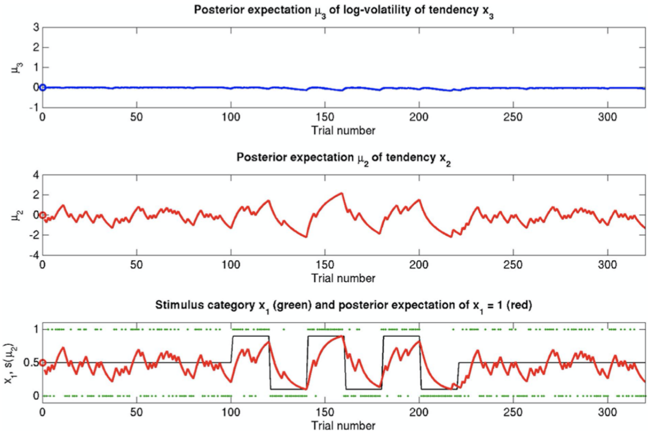

FIGURE 6 | Reduced $\vartheta = 0.05$ (unchanged $\omega = -2.2$, $\kappa = 1.4$). Symbols have the same meaning as in **Figure 5**. Here, the expected x_3 is more stable. The learning rate in x_2 is initially unaffected but owing to more stable x_3 it no longer increases after the period of increased volatility.

6.Simulations

Low sensory uncertainty

High sensory uncertainty

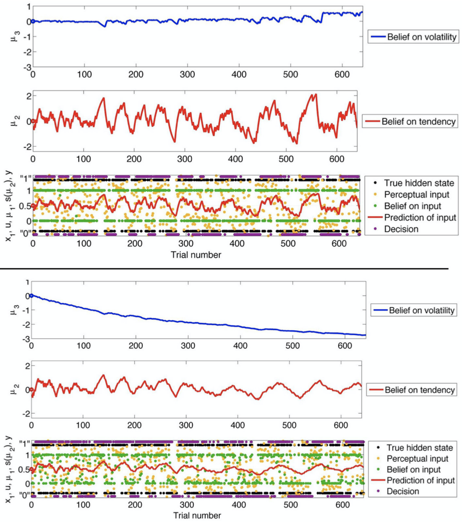

FIGURE 3 | The consequences of sensory uncertainty. Simulation of inference on a binary hidden state x_1 (black dots) using a three-level HGF under low ($\widehat{\pi}_u = 1000$, top panel) and high ($\widehat{\pi}_u = 10$, bottom panel) sensory uncertainty. Trajectories were simulated using the same input and parameters (except $\widehat{\pi}_u$) in both cases: $\mu_2^{(0)} = \mu_3^{(0)} = 0$, $\sigma_2^{(0)} = \sigma_3^{(0)} = 1$, $\kappa = 1$, $\omega = -3$, and $\vartheta = 0.7$. Decisions were simulated using a unit-square sigmoid model with $\zeta = 8$.